

4-14-2021

Personalized Game Recommendation Method Based on Implicit Feedback

Sha Jing

College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China;

Gongli Zeng

College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China;

Yang Yang

College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China;

Wei Yao

College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China;

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the [Artificial Intelligence and Robotics Commons](#), [Computer Engineering Commons](#), [Numerical Analysis and Scientific Computing Commons](#), [Operations Research](#), [Systems Engineering and Industrial Engineering Commons](#), and the [Systems Science Commons](#)

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

Personalized Game Recommendation Method Based on Implicit Feedback

Abstract

Abstract: Traditional recommendation systems often use explicit feedback for personalized recommendations. But the explicit feedback data is not easy to obtain, and the quality is poor, and the recommendation results unable to meet the requirement of the user. Implicit feedback data is easier to obtain and can provide users with the better content. *A personalized game recommendation method based on implicit feedback data is proposed. The method builds an implicit feedback recommendation model for game user data based on implicit feedback data such as the game duration and game numbers. A personalized recommendation of the game is implemented through an implicit semantic recommendation algorithm.* Through comparative experiments on a large number of real data sets, it is shown that the accuracy and recall of the proposed method are better than other methods.

Keywords

recommended system, implicit feedback, LFM(Latent Factor Model), personalized recommendations

Recommended Citation

Sha Jing, Zeng Gongli, Yang Yang, Wei Yao. Personalized Game Recommendation Method Based on Implicit Feedback[J]. Journal of System Simulation, 2021, 33(4): 809-817.

基于隐式反馈的个性化游戏推荐方法

沙静, 曾巩俐, 杨扬, 卫遥

(山东科技大学 计算机科学与工程学院, 山东 青岛 266590)

摘要: 传统的推荐系统常用显式反馈进行个性化推荐, 但显式反馈数据不易获取, 质量不好且易引起用户反感, 使推荐结果不能满足用户需求。隐式反馈数据更容易获取, 可更好地为用户提供其感兴趣的内容。提出一种基于隐式反馈数据的个性化游戏推荐方法。该方法基于游戏时长、游戏次数等隐式反馈数据, 构建针对游戏用户数据的隐式反馈推荐模型, 通过隐语义推荐算法实现了游戏的个性化推荐。通过大量真实数据集的对比实验, 表明该方法的精确率和召回率均优于其他方法。

关键词: 推荐系统; 隐式反馈; 隐语义模型; 个性化推荐

中图分类号: TH138; R318.6; TP391 文献标志码: A 文章编号: 1004-731X (2021) 04-0809-09

DOI: 10.16182/j.issn1004731x.joss.19-0636

Personalized Game Recommendation Method Based on Implicit Feedback

Sha Jing, Zeng Gongli, Yang Yang, Wei Yao

(College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China)

Abstract: Traditional recommendation systems often use explicit feedback for personalized recommendations. But the explicit feedback data is not easy to obtain, and the quality is poor, and the recommendation results unable to meet the requirement of the user. Implicit feedback data is easier to obtain and can provide users with the better content. A personalized game recommendation method based on implicit feedback data is proposed. The method builds an implicit feedback recommendation model for game user data based on implicit feedback data such as the game duration and game numbers. A personalized recommendation of the game is implemented through an implicit semantic recommendation algorithm. Through comparative experiments on a large number of real data sets, it is shown that the accuracy and recall of the proposed method are better than other methods.

Keywords: recommended system; implicit feedback; LFM(Latent Factor Model); personalized recommendations

引言

在互联网迅速发展的时代, 获取信息的速率得到了极大的提高, 但是随后带来了信息过载的问题。为了缓解信息过载问题, 使用户更加高效地从互联网获取所需信息, 推荐系统已经成为解决这些问题的主要工具。推荐系统通过分析用户的历史行为数据, 将信息进行科学的组织, 能够对每个不同的用户进行有效的个性化推荐, 给用户带来良好的

体验。

推荐算法主要分为3种: (1) 基于内容的推荐^[1]。根据用户和物品的特征数据, 为用户推荐可能会感兴趣的物品; (2) 协同过滤推荐^[2-3]。推荐领域中应用广泛的算法, 无需获得用户和物品的特征数据, 只需要用户的历史行为数据, 就可预测用户可能会感兴趣的物品; (3) 基于模型的协同过滤推荐^[4-5]。通过学习数据得出模型, 然后根据模型进行预测和推荐。

收稿日期: 2019-12-06 修回日期: 2020-04-13

基金项目: 国家自然科学基金(61170078, 61472228); 山东省科技发展计划(2012G0020120)

第一作者: 沙静(1976-), 女, 博士, 副教授, CCF会员, 研究方向为 Petri 网建模与验证, 系统性能分析, 大数据, 机器学习。E-mail: xo5547@163.com

用户历史行为数据从类型上主要分为显式反馈数据和隐式反馈数据。显式反馈数据是用户主动进行评价,例如评分、评级、评论等,能够明确地表达用户喜好的数据;隐式反馈数据则利用用户点击行为、用户浏览记录、购买历史等信息,不能直接表达用户喜欢与不喜欢,只是展现用户兴趣的数据^[6]。

用户评分等显式反馈虽然能够明显表达用户喜好,但受很多因素影响且不易获取,隐式反馈更易获取。隐式反馈存在于互联网上,它是用户在使用服务过程中留下的数据。用户在进行浏览新闻、购物、听音乐等行为中的各种选择,都可作为隐式反馈。

隐式反馈数据的收集范围广、成本低、不易引起用户反感,但数据规模一般比较大。隐式反馈虽然能够隐式地表达用户的兴趣和喜好,即能够表达用户的正倾向,却不易表达用户的负倾向,同时隐式反馈数据通常伴有噪声。

综上,如何高效处理大规模数据及合理有效地利用隐式反馈数据是传统推荐模型亟需解决的问题。本文提出了一种基于隐式反馈数据的游戏推荐方法,能够更准确地预测出游戏用户的兴趣偏好,同时在精确率和召回率上均优于对比算法。

1 相关工作

隐式反馈推荐不像显式反馈有明显的负反馈,隐式反馈可以很明确的知道用户喜欢什么,但却不能明确的知道用户不喜欢什么。Pan 等^[7]把基于隐式反馈数据的推荐问题定义为单类协同过滤问题,解决这个问题的主要思路是引入负样本:(1)人为指定负样本。通过人为设置正负样本的规则,从而指定负样本。例如 Jing 等^[8]在研究微博转发行为,设定用户在线浏览微博的过程,微博未被转发的为负样本,但是这个思路依赖于特殊的场景,不具有推广性;(2)从用户未选择的产品中进行随机抽样,抽样得到的样本定义为负样本^[7,9-10];(3)将所有用户没有选择过的产品都作为负样本,设置一个较小

的权重,权重反映了这些样本作为负样本的置信度^[8,11-12]。但在引入这些负样本的同时,也会引入噪声,增加训练的负担,影响训练效率,从而影响最终的推荐结果。

矩阵分解已经逐渐成为主流模型,渐渐地取代了依赖寻找相似用户^[13]或相似产品^[14]的基于近邻的协同过滤算法,矩阵分解具有可扩展性和灵活性,能够较准确地进行预测。最初的矩阵分解模型是奇异值分解(singular value decomposition, SVD)^[15],在此基础上,Koren 对 SVD 方法进行了改进,通过交替最小二乘法和随机梯度下降法提出融合用户、物品、偏置以及额外上、下文信息的 SVD++ 模型。Koren 在 SVD++ 的基础上进一步考虑时间问题,增加时序特征,提出 timeSVD++ 模型^[16],提高了模型预测的准确率。

但是矩阵分解本身并不适用于隐式反馈的场景,因为隐式反馈中缺少显式评分,因此在隐式反馈中一般是用 0-1 矩阵来描述用户-物品的行为,对于二值矩阵来说,一般 1 代表正反馈,即用户对该物品感兴趣,点击了或者购买了该物品,0 却不能完全认为是负反馈,0 一般有 2 种代表含义:用户对此物品不感兴趣,故没有点击;用户未见到此物品,因此没有点击。还有一种方法是将推荐任务转化为学习排序问题,即通过学习构造一个排序器将用户可能感兴趣的产品排在前面^[17]。Rendle 等^[18]提出贝叶斯个性化排序模型,此模型是将正例排在负例前面的概率进行最大化,然后给出协同排序的通用框架以及贝叶斯解释。

本文提出的模型是从数据集中选取负样本,并不需要再人为引入负样本,综合考虑了游戏与用户交互性强等因素,通过对用户偏好度和喜好度的计算等,准确预测出用户的兴趣所在。因此无论从模型角度还是实用性角度,本文的方法都能够较好地适用于隐式反馈的游戏推荐场景。

2 模型设计

本文研究的主要问题是基于隐式反馈的游戏

实时个性化推荐, 主要包括隐式反馈模型推荐和用户兴趣预测。流程图如图 1 所示。

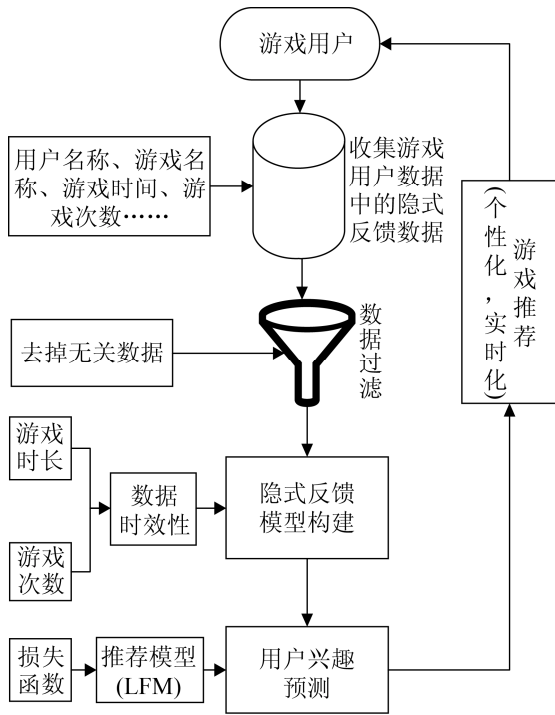


图 1 基于隐式反馈数据的游戏实时个性化推荐
Fig. 1 Real-time personalized recommendations for games based on implicit feedback data

首先收集游戏用户数据中的隐式反馈数据, 数据中包括用户名称、游戏名称、游戏时间、游戏次数等内容。收集的数据需要经过数据过滤, 常用的数据过滤方法有比较运算, 范围运算, 空值匹配, 字符匹配和逻辑匹配, 本文运用范围运算和空值匹配方法去掉与本实验无关的数据, 这样可以减小数据的运算量并增加数据的有效性。然后通过偏好值的假设、计算喜好程度等构建隐式反馈推荐模型, 最后在已有数据集的基础上, 用隐语义模型 (Latent Factor Model, LFM) 算法^[19-21]实现对游戏用户的兴趣预测, 为游戏用户推荐他们感兴趣的内容, 实现个性化推荐。

2.1 隐式反馈推荐模型

目前大多数评分推荐系统都是依据显式评分而来的, 而且默认数据来源是可靠的。但是显式评分并不总是可得, 影响因素比较多, 容易引起用户

反感, 因此较多沉默用户会选择忽略评分。

由于显式反馈的原始数据本身质量就存在问题, 在这些数据的基础上进行推荐, 推荐结果的有效性和准确性并不高。

为了给用户带来更好的游戏体验, 相对于显式反馈数据推荐, 隐式反馈数据推荐的内容更丰富、更准确有效, 不容易受到用户偏见的影响, 所以数据更具真实性、有效性, 因此能够更好地体现游戏用户的真实喜好^[22]。但是隐式反馈数据中只有正样本, 负样本一般需要人为引入, 所以显式反馈数据的推荐模型不一定适用于隐式反馈的推荐模型。根据隐式反馈的数据的特性并结合显式反馈数据的推荐模型, 构造隐式反馈的推荐模型。

显式反馈数值可以代表用户对物品的偏好程度, 而隐式反馈的数值通常是代表动作的频次, 频次越多并不代表偏好值越大。本文定义: 用户 i 和项目 j , 在显式反馈中 a_{ij} 一般表示用户 i 对项目 j 的偏好值, 在隐式反馈中 a_{ij} 为动作频次(本文即游戏点击次数), h_{ij} 为隐式反馈中的偏好值, 对于偏好值的假设:

$$h_{ij} = \begin{cases} 1, a_{ij} > 0 \\ 0, a_{ij} = 0 \end{cases} \quad (1)$$

用户 i 对物品 j 的动作频次超过一次, 就认为偏好值为 1, 用户 i 喜欢物品 j 。用户 i 对物品 j 的动作频次小于 1 次, 就认为偏好值为 0, 假设的置信度就会很低。用户没有点击此物品, 可能是不喜欢、不感兴趣, 但也有其他原因, 可能是没见过或者不知道它的存在等, 因此不能确定用户是否喜欢该物品。在显式反馈中的矩阵分解优化目标函数一般通过式(2)来表达:

$$(\hat{U}, \hat{V}) = \arg \min_{U, V} \sum_{(u, i) \in A} (M_{ui} - [UV^T]_{ui})^2 \quad (2)$$

式中: $U \in \mathbb{R}^{m \times r}$; $V \in \mathbb{R}^{m \times r}$ ($r \ll m, n$); $M \in \mathbb{R}^{m \times n}$, 其中 $\mathbb{R}^{m \times n}$ 是 m 个用户和 n 个项目的评分矩阵; A 是从矩阵 M 内已知评分中选取的训练集合。

通过式(2)的设计思路可以类比得到本文的优化目标函数, 结合游戏推荐背景, 增加了游戏用户

对游戏的操作时长等因素,为了使其能够更适应隐式反馈的特点,设置了喜好程度和偏好值等。最终本文的优化目标函数可通过式(3)描述:

$$r_{ui} = \min \sum_{i=1}^m \sum_{j=1}^n c_{ij} (h_{ij} - u_i v_j^T)^2 + d_{jk}^i (T_0^i - t_{jk}^u) \quad (3)$$

式中: c_{ij} 为用户 i 对游戏 j 的喜好程度; h_{ij} 为隐式反馈中的偏好值; u_i 和 v_j 为用户 i 和游戏项目 j 的特征值; d_{jk}^i 为游戏用户操作游戏 j 的第 k 条记录的操作时长; T_0^i 为 k 天集合内用户 i 的所有游戏操作记录中的最后一次操作开始时间; t_{jk}^u 为游戏用户操作游戏 j 的第 k 条记录的开始时间。

随着用户点击游戏次数 a_{ij} 的增加,表示用户 i 对游戏 j 的喜好程度增加,因此本文用 c_{ij} 表示用户 i 对游戏 j 的喜好程度(置信度)。通过式(4)表示:

$$c_{ij} = 1 + \omega a_{ij} \quad (4)$$

式中: a_{ij} 为用户点击游戏的次数; ω 为喜好程度的计算系数。当 $\omega=30$ 时,能获得最优值。由式(4)可以看出隐式反馈中点击游戏频次越多, c_{ij} 的值越大。

本文的隐式反馈推荐是针对所有的 i, j 键值对,而不像显式反馈只针对可观察到的数据。所以隐式反馈中一共有 $m \times n$ 个数据,对于没有反馈的项,本文赋予较小的置信度。

反馈模型是通过对数据的累加得到的,而且在计算的过程充分考虑了游戏的操作次数和游戏时长,因此因玩家误操作对结果造成的影响比较小。通过加入喜好程度的计算,可以避免模型发生偏斜,同时可避免引入噪声。通过目标函数的计算能够得到理想状态下的结果,在接下来的用户兴趣预测部分提供预测标准。

2.2 用户兴趣预测

在实际应用中,很多显式反馈推荐算法并不能直接应用于隐式反馈推荐中,所以本文用隐语义模型(LFM)方法实现对游戏用户的兴趣预测。LFM 是通过隐含特征联系用户兴趣和物品。假设,已知一个关于某购物网站的用户-购买物品评

分的数据集 D ,在数据集 D 中,购物网站的用户个数为 U ,购买物品个数为 T , D 是 $U \times T$ 的矩阵^[23]。通过 LFM 方法建模后,得到如图 2 所示的隐语义模型。



图2 隐语义模型矩阵分解示意图

Fig. 2 Schematic diagram of matrix decomposition for implicit semantic model

矩阵 M 是 *user* (用户)对 *factor* (因子类别)的兴趣度,矩阵 N 是 *item* (物品)在 *factor* (因子类别)上的权值,两者的内积是用户对物品的总体兴趣度值。

因此,本文对用户兴趣的预测,首先需要用户对用户的兴趣进行分类,再从分类的结果中挑选出用户可能会感兴趣的物品。本文采用隐含语义分析技术^[21]对用户兴趣进行分类。

隐含语义分析技术是采用基于用户行为统计的自动聚类,它在物品分类上与 ItemCF 在物品分类的思想相似。假设同时有 2 个物品被许多用户所喜欢,那么这 2 个物品就有很大的概率属于同一个类型;隐含语义分析技术中每个类都具有不同的维度,它是通过计算用户所拥有的共同兴趣而得到的,假设用户的共同兴趣是一个维度 A ,那么 LFM 所得出的类也是维度 A ;隐含语义分析技术会计算出物品属于某个类的权重,所以物品是通过计算而合理地进行分类。用户的行为决定每个类中物品的权重,如果喜欢某个类的用户可能都会喜欢这个类中的某一个物品,那么这个物品在这个类中占有的权重就会比较高^[24]。

LFM 通过式(5)计算游戏用户 u 对物品 i 的兴趣, p_{uk} , q_{ik} 为模型的参数, p_{uk} 是计算游戏用户 u 的兴趣和第 k 个隐类之间的关系,而 q_{ik} 则是计算物品 i 和第 k 个隐类之间的关系,公式为:

$$Preference(u, i) = p_u^T q_i = \sum_{f=1}^F p_{uf} q_{if} \quad (5)$$

在隐式反馈的数据集上应用 LFM 进行推荐, 首先是如何给每个游戏用户生成负样本。对于生成负样本, 有 2 个原则: (1) 要保证正负样本的平衡即游戏数据集中正负样本的数目要基本相同; (2) 对于游戏用户选取的负样本, 要尽量选取游戏用户没有进行点击行为的热门物品。因为热门物品相对于冷门物品来说, 用户见到的可能性要比较大, 所以用户对于热门物品没有点击行为的, 就可能代表用户对此物品不感兴趣。对于冷门物品。用户没有见到过的可能性会更大, 所以即使用户没有对物品进行点击行为, 也并不代表用户对物品不感兴趣。

因此, 需要优化式(6)所示的损失函数, 找到最合适的参数 p 和 q :

$$C = \sum_{(u,i) \in K} (r_{ui} - \hat{r}_{ui})^2 = \sum_{(u,i) \in K} \left[\left(r_{ui} - \sum_{k=1}^K p_{uk} q_{ik} \right)^2 + \lambda \|p_u\|^2 + \lambda \|q_i\|^2 \right] \quad (6)$$

其中, 正则化参数是为了避免模型过度拟合, 即 $\lambda \|p_u\|^2 + \lambda \|q_i\|^2$ 为防止过拟合项。

为了降低参数更新时的计算量, 最小化式(6)的损失函数, 本文采用随机梯度下降法(Stochastic Gradient Descent, SGD)^[19]。通过求参数的偏导数找到最速下降的方向, 其中 batch size=30:

$$\frac{\partial C}{\partial p_{uk}} = -2q_{ik} \cdot e_{ui} + 2\lambda p_{uk} \quad (7)$$

$$\frac{\partial C}{\partial q_{ik}} = -2p_{uk} \cdot e_{ui} + 2\lambda q_{ik} \quad (8)$$

其中 $e_{ui} = r_{ui} - \sum_{k=1}^K p_{uk} q_{ik}$ (9)

然后将参数沿着最速下降的方向继续前进, 通过迭代的方法不断优化参数:

$$p_{uk} = p_{uk} + \alpha(q_{ik} \cdot e_{ui} - \lambda p_{uk}) \quad (10)$$

$$q_{ik} = q_{ik} + \alpha(p_{uk} \cdot e_{ui} - \lambda q_{ik}) \quad (11)$$

其中正则化参数 λ 、学习率 α 可以通过反复实验来获得。

3 实验评估

3.1 数据集与评价标准

本文实验的数据集采用的是某网吧平台运营商的游戏数据集记录(游戏时长, 游戏次数, 点击记录, 购买记录等), 时间跨度由 2018-07-01—2018-09-01。数据集经过匿名处理和清除与游戏无关的数据、噪声数据后, 获得 15 886 724 条游戏操作记录。

在一般的预测准确度的标准中, 评分预测通常是用均方根误差(RMSE)和平均绝对误差(MAE)来衡量的, 其计算方法如式(12), (13)所示:

$$RMSE = \sqrt{\frac{\sum_{u,i \in T} (r_{ui} - \hat{r}_{ui})^2}{|T|}} \quad (12)$$

$$MAE = \frac{\sum_{u,i \in T} |r_{ui} - \hat{r}_{ui}|}{|T|} \quad (13)$$

式中: r_{ui} 为用户 u 对物品 i 的实际评分; \hat{r}_{ui} 为推荐算法给出的预测评分; T 为观测次数。

然而这 2 个并不适用于基于隐式反馈的推荐中, 因为 r_{ui} 是用户给出的对物品的实际评分, 而在隐式反馈中并没有实际的显式评分, 所以无法计算 RMSE 和 MAE。而本文的目的是对游戏用户进行个性化推荐, 注重的是最终的推荐结果, 而不是预测的评分与实际评分的误差大小。

因此本文实验采用精确率(precision)和召回率(recall)作为实验的评估标准, 精确率表示预测正确的样本总数占总样本数的比例, 召回率表示测试集中推荐列表命中的物品数和用户点击的物品的集合大小的比例, 召回率越大, 效果越好。其计算方法如式(14), (15)所示:

$$precision = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |R(u)|} \quad (14)$$

$$recall = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |T(u)|} \quad (15)$$

式中: $R(u)$ 为本实验根据游戏用户在训练集上的行为所做出的推荐列表集合; $T(u)$ 是游戏用户在测试

机上实际喜欢游戏的集合。

3.2 对比算法

基于隐式反馈的游戏推荐算法很少,同时考虑隐式反馈推荐应用到游戏场景的特点以及数据集本身的特点,一些已有的隐式反馈推荐算法并不适用于游戏场景,所以本文选择 5 种算法作为对比算法。为了保证实验效果的有效性和说服力,对比算法都采用与本文实验的游戏数据集,实验中所涉及的参数均为最优参数:

(1) 基于物品流行度的推荐(Popular): 该算法按照物品的流行度进行推荐,在本实验中游戏的流行度按照游戏对应的玩家人数确定;

(2) 基于用户的协同过滤推荐(User-based CF): 该算法首先根据数据集计算与用户有相似兴趣的其他用户集合,然后推荐用户喜欢但用户没有听说过的集合中的物品^[25];

(3) 基于物品的协同过滤推荐(Item-based CF): 该算法首先根据数据集为每个产品计算出相似的产品,然后给用户推荐和他们之前喜欢的物品相似的物品;

(4) 贝叶斯个性化排序(Bayesian Personalized Ranking, BPR)^[26]: 该算法将推荐问题转换为成对排序的方法,从每个正例中随机抽样得到负例形成对,然后推荐商品对应喜好比较靠前的物品。该方法也常应用于隐式反馈的推荐场景中;

(5) SVD^[27]: 该算法采用随机梯度下降法优化初始化的用户特征矩阵和物品特征矩阵,使其接近原始的评分矩阵,常应用于隐式反馈推荐中。

3.3 精确率和召回率

该实验通过将 LFM 算法与其他 5 种算法作对比,为了获得更加可靠的实验结果,本实验选取了 70% 的数据作为训练集,选取 30% 作为测试集,并将实验重复构造 3 次,取 3 次实验结果的平均值作为最后的结果,根据一般通用设置,设置的正则化参数 $\lambda=0.05$,为达到理想推荐效果,模型的迭代次数为 40 次。

在评估标准中本文引入数据集中的时间跨度 T , 评估在不同时间跨度 T 下实验的精确率和召回率的变化情况。图 3~4 为不同时间跨度 T 各算法的精确率和召回率的对比。

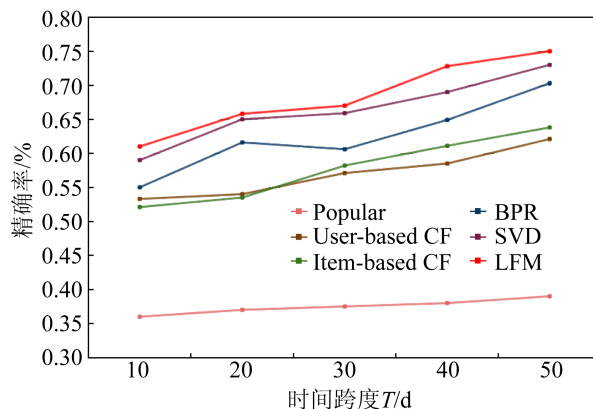


图 3 不同时间跨度 T 下精确率对比

Fig. 3 Comparison of accuracy at different time spans T

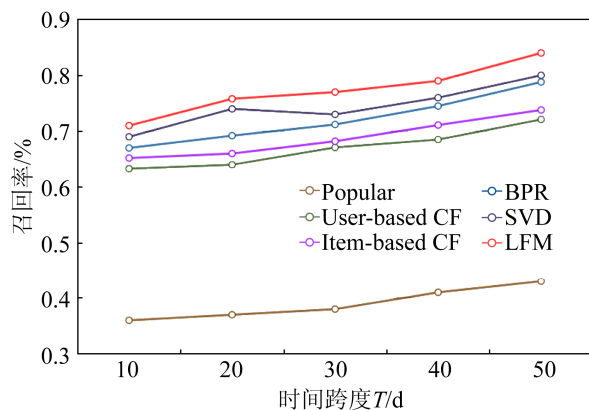


图 4 不同时间跨度 T 下召回率对比

Fig. 4 Comparison of recall rate under different time span T

从图 3 可以看出,本文所引用的 LFM 在不同时间跨度 T 下,它的精确率都高于其他对比算法;其中 Popular 算法的精确率最低,比 User-based CF、Item-based CF 都要低,这说明了基于流行度的计算方法的精确率是要低于基于矩阵分解的算法的。从图 3 能够看出随着时间跨度的增加即数据集的增加,算法的精确率趋势都趋于平缓,这表明数据集越多,预测的结果越准确越稳定,所以可以根据实际推荐需求选取合适的时间跨度 T 即合适的数据集,从而选择最合适的计算量,避免进行不必要的计算,以最优的速度获得推荐结果。

从图 4 可以看出, 总体趋势与图 3 大致相同。召回率越大, 证明算法的实验效果越好。所以从图 4 可以得出, 本文所采用 LFM 方法的召回率最高, 效果最好。LFM 结合用户游戏时长和游戏次数对用户进行推荐, 比只根据流行度或者游戏时长或游戏次数的推荐效果要好, 命中率要高, 同时 LFM 重视数据的时效性并可较好地实现在线推荐。通过实验可以看出用户对游戏的游戏偏好程度并不能只单单用游戏时长和游戏次数其中的一个决定, 同时也说明了数据的时效性对推荐结果的精确度等都有影响。仅仅只根据游戏的流行度对用户进行推荐的效果最差, 所以在隐式推荐中应该综合考虑多种因素, 从而实现对用户的个性化推荐, 让大部分用户对推荐结果满意。

3.4 实验参数对精确率和召回率的影响

在 LFM 中, 重要的参数有 4 个: 隐形特征个数 F , 学习速率 α , 正则化参数 λ , 负样本比例 $ratio$ 。其中正则化参数 λ 是为了避免模型过度拟合, 控制模型与真实场景的拟合程度。图 5 为不同 λ 的取值对精确率和召回率的影响。

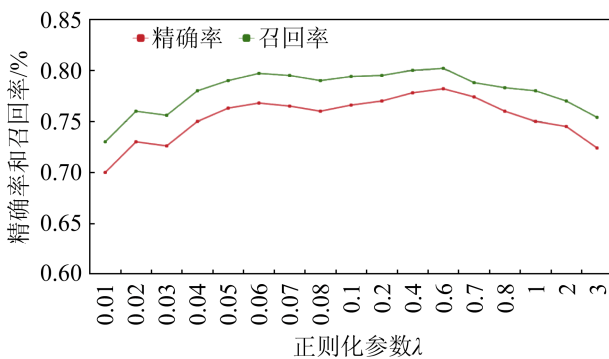


图 5 精确率和召回率在不同 λ 值的变化
Fig. 5 Changes in accuracy and recall with different values of λ

通过实验结果可以看出, λ 在 0.08~0.60 之间的取值精确率和召回率的效果是最好的, λ 的取值太小会造成模型过度拟合, 太大会造成模型不能很好的拟合。所以经过实验可以选取合适的 λ 值, 使得实验效果更好。

图 6 为在不同的隐性特征个数 F 的取值下精确率和召回率的变化。

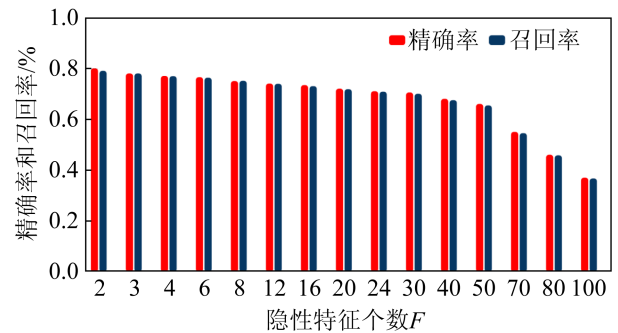


图 6 精确率和召回率在不同 F 值的变化
Fig. 6 Changes in accuracy and recall with different values of F

由图 6 可以看出, 随着隐性特征个数 F 的增加, 推荐结果的精确率和召回率呈下降趋势, 这说明对于本文研究的游戏来说, 它的隐形特征个数相对来说比较单一, 主要是依靠几个关键的特征来吸引游戏用户, 并不是隐形特征越多越好。

图 7 为不同 $ratio$ 参数(负样本的比例)的取值对推荐结果精确率和召回率的影响。

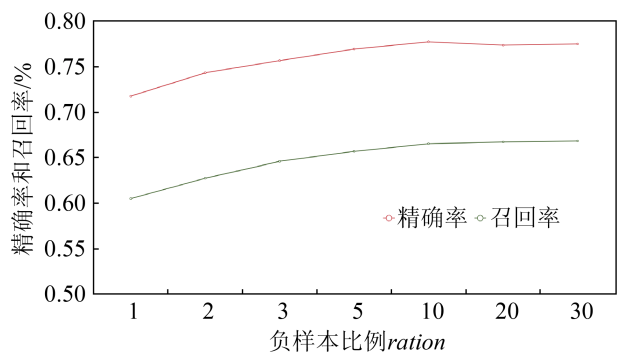


图 7 精确率和召回率在不同 $ratio$ 值的变化
Fig. 7 Changes in accuracy and recall with different values of $ratio$

通过实验发现, $ratio$ 参数对 LFM 的性能影响最大。由图 7 可以看出, 当 $ratio \leq 10$ 时, 精确率和召回率会随着 $ratio$ 的增加有显著的提高。当 $ratio > 10$ 后, 图像趋于平缓, 即精确率和召回率基本保持稳定。

在梯度下降法中, 学习率 α 是一个比较重要的

参数, 它会影响收敛的速度和最后的优化效果。经过实验得到: 当初始学习率 $\alpha=0.02$ 时, 经过 3 次迭代的推荐结果就已经达到峰值并且无法继续提升, 且后续的迭代中逐渐出现推荐效果下降的趋势; 当初始学习率 $\alpha=0.001$ 时, 可以达到理想的推荐效果, 但需要 40 次的迭代, 这说明当学习率过小的时候会导致收敛速度过慢。所以还需要进一步的实验来动态调整学习率, 使得迭代次数也调整为合适的值。

4 结论

本文提出了基于隐式反馈数据的对游戏用户的个性化推荐方法。通过 LFM 的方法对用户的行为进行分析, 建模实验等处理, 实现对用户的个性化推荐, 使得用户最大限度的获得满意的推荐效果。在大量真实数据集上的实验验证了本文方法的有效性。

未来将在此模型的基础上进一步研究数据更新时的推荐结果动态实时更新问题, 以及当数据量更大(TB 以上)时的计算性能问题。另外, 本文是基于纯隐式反馈数据上的推荐, 并没有考虑显式反馈数据的问题。在实际生活场景中, 显式反馈和隐式反馈是可能并存的, 所以我们接下来将研究这两者能否更好的结合, 以提高推荐效果。

参考文献:

- [1] Pazzani M J, Billsus D. Content-based Recommendation Systems[M]. The Adaptive Web. Berlin: Springer, 2007: 325-341.
- [2] 李伟霖, 王成良, 文俊浩. 基于评论与评分的协同过滤算法[J]. 计算机应用研究, 2017, 34(2): 361-364.
Li Weilin, Wang Chengliang, Wen Junhao. Reliability Analysis of Manipulator based on Fourth-Momentestimation[J]. Application Research of Computers, 2017, 34(2): 361-364.
- [3] Wang C D, Deng Z H, Lai J H, et al. Serendipitous Recommendation in E-commerce Using Innovator-based Collaborative Filtering[J]. IEEE transactions on cybernetics (S2168-2267), 2019, 49(7): 2678-2692.
- [4] 李博, 陈志刚, 黄瑞, 等. 基于 LDA 模型的音乐推荐算法[J]. 计算机工程, 2016, 42(6): 175-179.
Li Bo, Chen Zhigang, Huang Rui, et al. Music Recommendation Algorithm based on LDA Model[J]. Computer Engineering, 2016, 42(6): 175-179.
- [5] Chen C, Zheng X, Wang Y, et al. Context-aware Collaborative Topic Regression with Social Matrix Factorization for Recommender Systems[C]// 28th AAAI Conference on Artificial Intelligence. Québec: AAAI Press, 2014: 9-15.
- [6] Deng Z H, Huang L, Wang C D, et al. DeepCF: a unified Framework of Representation Learning and Matching Function Learning in Recommender System[C]// 33rd AAAI Conf. on Artificial Intelligence. Palo Alto: AAAI Press, 2019: 61-68.
- [7] Pan R, Zhou Y, Cao B, et al. One-class Collaborative Filtering[C]// 8th IEEE International Conference on Data Minin. Pisa: IEEE, 2008: 502-511.
- [8] Jiang M, Cui P, Liu R, et al. Social Contextual Recommendation[C]// 21st ACM Int'l Conf. on Information and Knowledge Management. Maui, Hawaii, USA: ACM Press, 2012: 45-54. [doi: 10.1145/2396761.2396771]
- [9] Chen K, Chen T, Zheng G, et al. Collaborative Personalized Tweet Recommendation[C]//35th Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. Portland Oregon USA: ACM Press, 2012: 661-670. [doi: 10.1145/2348283. 2348372]
- [10] Shi Y, Karatzoglou A, Baltrunas L, et al. TFMAP: Optimizing MAP for Top-n Context-aware Recommendation[C]//35th Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. Portland Oregon USA: ACM Press, 2012: 155-164. [doi: 10.1145/2348283. 2348308]
- [11] Hu Y, Koren Y, Volinsky C. Collaborative Filtering for Implicit Feedback Datasets[C]//8th IEEE Int'l Conf. on Data Mining (ICDM 2008). Pisa, Italy: IEEE, 2008: 263-272. [doi: 10.1109/ICDM.2008.22]
- [12] Steck H. Training and Testing of Recommender Systems on Data Missing not at Random[C]//16th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. Washington DC USA: ACM Press, 2010: 713-722. [doi: 10.1145/1835804. 1835895]
- [13] Herlocker J L, Konstan J A, Borchers A, et al. An Algorithmic Framework for Performing Collaborative Filtering[C]//22nd Annual Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. Berkeley California USA: ACM Press, 1999: 230-237. [doi:10.1145/312624. 312682]

- [14] Sarwar B, Karypis G, Konstan J, et al. Item-Based Collaborative Filtering Recommendation Algorithms[C]// 10th Int'l Conf. on World Wide Web. Hong Kong: ACM Press, 2001: 285-295. [doi: 10.1145/371920.372071]
- [15] Paterek A. Improving Regularized Singular Value Decomposition for Collaborative Filtering[C]//KDD Cup and Workshop. San Jose: ACM Press, 2007: 5-8.
- [16] Koren Y. Collaborative Filtering with Temporal Dynamics[J]. Communications of the ACM (S0001-0782), 2010, 53(4): 89-97.
- [17] 印鉴, 王智圣, 李琪, 等. 基于大规模隐式反馈的个性化推荐[J]. 软件学报, 2014, 25(9): 1953-1966.
Yin Jian, Wang Zhisheng, Li Qi, et al. Personalized Recommendation based on Large-scale Implicit Feedback[J]. Journal of Software, 2014, 25(9): 1953-1966.
- [18] Rendle S, Freudenthaler C, Gantner Z, et al. BPR: Bayesian Personalized Ranking from Implicit Feedback[C]//25th Conf. on Uncertainty in Artificial Intelligence. Montreal, Canada: AUAI Press, 2009: 452-461.
- [19] 吴宾, 娄铮铮, 叶阳东. 联合正则化的矩阵分解推荐算法[J]. 软件学报, 2018, 29(9): 2681-2696.
Wu Bin, Lou Zhengzheng, Ye Yangdong. Co-regularized Matrix Factorization Recommendation Algorithm[J]. Journal of Software, 2018, 29(9): 2681-2696.
- [20] 刘华锋, 景丽萍, 于剑. 融合社交信息的矩阵分解推荐方法研究综述[J]. 软件学报, 2018, 29(2): 340-362.
Liu Huafeng, Jing Liping, Yu Jian. Survey of Matrix Factorization based Recommendation Methods by Integrating Social Information[J]. Journal of Software, 2018, 29(2): 340-362.
- [21] Lin C Y, Wang L C, Tsai K H. Hybrid Real-time Matrix Factorization for Implicit Feedback Recommendation Systems[J]. IEEE Access (S2169-3536), 2018, 6: 21369-21380.
- [22] 俞东进, 陈聪, 吴建华, 等. 基于隐式反馈数据的个性化游戏推荐[J]. 电子学报, 2018, 46(11): 2626-2632.
Yu Dongjin, Chen Cong, Wu Jianhua, et al. Personalized Game Recommendation Based on Implicit Feedback[J]. Electronic Journal, 2018, 46(11): 2626-2632.
- [23] 李琳娜, 江雪琴. 推荐系统中的隐语义模型研究[J]. 情报工程, 2016, 2(4): 30-39.
Li Linna, Jiang Xueqin. Research on the Latent Factor Model in Recommender System[J]. Intelligence engineering, 2016, 2(4): 30-39.
- [24] 邵波. 基于信息匹配的推荐算法研究[D]. 沈阳: 沈阳建筑大学, 2015.
Shao Bo. Research on Recommendation Algorithm based on Information Matching [D]. Shenyang: Shenyang Jianzhu University, 2015.
- [25] 王俊淑, 张国明, 胡斌. 基于深度学习的推荐算法研究综述[J]. 南京师范大学学报(工程技术版), 2018, 18(4): 33-43.
Wang Junshu, Zhang Guoming, Hu Bin. A Review of Research on Recommendation Algorithms Based on Deep Learning[J]. Journal of Nanjing Normal University (Engineering and Technology Edition), 2018, 18(4): 33-43.
- [26] 程明月, 刘淇, 李徽, 等. 多重对级贝叶斯个性化排序算法[J]. 南京信息工程大学学报(自然科学版), 2019, 11(3): 302-308.
Cheng Mingyue, Liu Qi, Li Hui, et al. Multi-pair Bayesian Personalized Ranking[J]. Journal of Nanjing University of Information Science and Technology (Natural Science Edition), 2019, 11(3): 302-308.
- [27] Li G, Chen Q. Exploiting Explicit and Implicit Feedback for Personalized Ranking[J]. Mathematical Problems in Engineering (S1024-123X), 2016(1): 1-11.